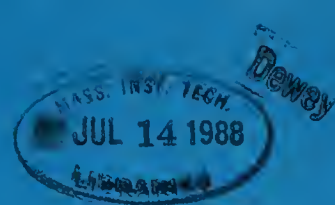


Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

<http://www.archive.org/details/specificationtes00wool>

HB31
.M415
no. 479



**working paper
department
of economics**

SPECIFICATION TESTING AND
QUASI-MAXIMUM LIKELIHOOD ESTIMATION

by

Jeffrey M. Wooldridge

No. 479

December 1987

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**



SPECIFICATION TESTING AND
QUASI-MAXIMUM LIKELIHOOD ESTIMATION

by

Jeffrey M. Wooldridge

No. 479

December 1987

M.I.T. LIBRARIES
JUL 15 1988
RECEIVED

SPECIFICATION TESTING AND QUASI-MAXIMUM LIKELIHOOD ESTIMATION

Jeffrey M. Wooldridge
Department of Economics
Massachusetts Institute of Technology, E52-262C
Cambridge, MA 02139
(617) 253-3488

February 1987
Revised: December 1987

I wish to thank Daniel McFadden, Danny Quah, Jerry Hausman, and the participants of the Harvard/MIT and Yale econometrics workshops for helpful comments. I am responsible for any errors.

ABSTRACT: This paper develops robust, regression-based forms of Newey's conditional moment tests for models estimated by quasi-maximum likelihood using a density in the linear exponential family. A novel feature of these tests is that, in addition to the original estimation, they require only two linear least squares regressions for computation, while remaining robust to distributional misspecifications other than those being explicitly tested. Several examples are presented to illustrate the simplicity and scope of the procedure: a Lagrange multiplier test for nonlinear regression, the score form of the Hausman test for the parameters of a conditional mean, and a regression form of the Davidson-MacKinnon nonnested hypotheses test. All of the tests assume only that the conditional mean is correctly specified under the null hypothesis.

Tests for second moment misspecification, developed using White's information matrix testing principle, assume only that the first two moments are correctly specified under the null hypothesis. A special case is a regression-based test for heteroskedasticity in nonlinear models which relaxes the assumption that the conditional fourth moment of the errors is constant. Also, a simple distributional test for the Poisson regression model is presented.

KEYWORDS: Conditional moment tests, robustness, quasi-maximum likelihood, linear exponential family.

1. INTRODUCTION

Many economic hypotheses can be formulated in terms of the conditional expectation $E(Y_t|X_t)$ of one set of variables Y_t given a set of predetermined variables X_t . If the conditional expectation is known up to a finite number of parameters then hypotheses of interest can be formulated as restrictions on parameters; classical inference procedures are then available for formally carrying out the appropriate tests.

Sometimes economists are interested in comparing two models for Y_t , neither of which contains the other as a special case. In this case the competing economic hypotheses are nonnested in a statistical sense, and classical testing procedures (e.g. Wald, Likelihood Ratio and Lagrange Multiplier tests) are no longer applicable. There are, however, several tests available in the presence of nonnested alternatives. The Cox (1961,1962) approach is useful in the general maximum likelihood setting. Davidson and MacKinnon (1981) derive tests for nonnested regression models.

If one is not interested in specific alternatives to the postulated regression function, but is concerned about misspecification which leads to inconsistent estimates of economic parameters, then the Hausman (1978) methodology is available.

This paper develops a class of Newey's (1985) conditional moment (CM) tests that are explicitly designed to detect misspecification of a conditional expectation. The class of tests considered is broad enough to contain the three types of tests mentioned above, and allows for time series as well as cross section

observations.

Broadly speaking, the setup here is encompassed by White (1985b), who extends Newey's work and develops a framework which includes conditional moment tests for time series observations as a special case. However, specializing White's results to a class of moment restrictions intended to detect (dynamic) misspecification of the regression function, and restricting attention to the class of quasi-maximum likelihood estimators (QMLE's) derived from a density in the linear exponential family (LEF), allows derivation of simple forms of these tests without additionally assuming that the conditional density is correct under the null hypothesis. If the conditional mean is the object of interest, then a test which further assumes that the distribution is correctly specified will generally have the wrong asymptotic size for testing the relevant null hypothesis. Moreover, standard regression forms of conditional mean tests are inconsistent for testing distributional assumptions beyond the first moment.

A useful feature of the robust tests derived here is that, in addition to the QMLE estimation, only two linear least squares regressions are required for computation. In many cases the statistics needed for the auxiliary regressions are computable from the final iteration of the Berndt, Hall, Hall and Hausman (BHHH) algorithm. One simple consequence of the results here is that heteroskedasticity-robust Lagrange Multiplier tests for exclusion restrictions in a dynamic linear model are computable by running a total of three linear regressions.

Although correctly specifying the regression function is usually the primary concern of the applied econometrician, it is also useful to know whether the distribution is correctly specified. Section 5 of this paper derives a modification of White's (1982) information matrix test which is designed to detect misspecification of the conditional second moment. An interesting feature of the test derived here is that it is regression-based while remaining robust to misspecification of other aspects of the distribution: the null hypothesis states only that the first two conditional moments are correctly specified. As a special case, it yields a regression form of the White (1980a) test for heteroskedasticity for nonlinear regression which does not require that the errors have a constant conditional fourth moment. In addition, it gives simple tests for distributional specification in such interesting cases as the Poisson regression model.

The remainder of the paper is organized as follows. Section 2 introduces the general setup and briefly discusses some useful properties of LEF distributions. Section 3 derives the computationally simple conditional mean test statistics. Several examples of conditional mean tests are presented in Section 4. Section 5 considers a modified information matrix test for conditional second moments, and Section 6 contains some concluding remarks.

2. NOTATION AND SETUP

Let $\{(Y_t, Z_t): t=1,2,\dots\}$ be a sequence of observable random vectors with Y_t $1 \times K$, Z_t $1 \times L$. The variables Y_t are the dependent or endogenous variables. Interest lies in explaining Y_t in terms of the explanatory variables Z_t and (in a time series context) past values of Y_t and Z_t . Let $X_t \equiv (Z_t, Y_{t-1}, Z_{t-1}, \dots, Y_1, Z_1)$ denote the predetermined variables (Z_t may be excluded from X_t without alternating the following results). The conditional distribution of Y_t given $X_t = x_t$ always exists and is denoted $D_t(\cdot | x_t)$. Under weak conditions on $D_t(\cdot | x_t)$ there exists a conditional density $p_t(y_t | x_t)$ with respect to a σ -finite measure $\nu_t(dy_t)$ (see Wooldridge (1987)). Because (by definition) we are not interested in the stochastic behavior of $\{Z_t\}$, the conditional densities $p_t(y_t | x_t)$ describe the relevant dynamic behavior of $\{Y_t\}$.

We choose for $p_t(y_t | x_t)$ a class of conditional densities $\{f_t(y_t | x_t, m_t, \eta_t)\}$ (which may or may not contain $p_t(y_t | x_t)$) which comprises of members of the linear exponential family. In particular,

$$(2.1) \quad \log f_t(y_t | x_t, m_t, \eta_t) = \\ a(m_t(x_t), \eta_t(x_t)) + b(y_t, \eta_t(x_t)) + y_t c(m_t(x_t), \eta_t(x_t))$$

where m_t is $1 \times K$, η_t is $1 \times J$, $c(\cdot)$ is $K \times 1$, and $a(\cdot)$ and $b(\cdot)$ are scalars. The function $m_t(x_t)$ is the expectation associated with the density $f_t(y_t | x_t, m_t, \eta_t)$. This entails the restriction

$$(2.2) \quad m_t(x_t) \nabla_m c(m(x_t), \eta_t(x_t)) = -\nabla_m a(m_t(x_t), \eta_t(x_t))$$

for all x_t .

Following Gourieroux, Monfort and Trognon (1984a) (hereafter GMT (1984a)), the LEF family is parameterized through the conditional mean function

$$(2.3) \quad \{m_t(x_t, \theta) : \theta \in \Theta \subset \mathbb{R}^P\}$$

and the nuisance parameters

$$(2.4) \quad \{n_t(x_t, \pi) : \pi \in \Pi \subset \mathbb{R}^N\}.$$

In the context of Section 3, correct specification means correct (dynamic) specification of the conditional mean, i.e.

$$(2.5) \quad E(Y_t | X_t = x_t) = m_t(x_t, \theta_0) \quad \text{for some } \theta_0 \in \Theta, t=1,2,\dots$$

As GMT (1984a) have shown in the case of independent observations and as White (1985a) has shown in a more general dynamic setting, the LEF class of densities has the useful property of consistently estimating the parameters of a correctly specified conditional expectation despite misspecification of other aspects of the conditional distribution.

The nuisance parameter π may be assigned any value in Π ; more generally, it may be replaced by an estimator $\hat{\pi}_T - \pi_T^0 \xrightarrow{P} 0$ where $\{\pi_T^0\} \subset \Pi$. The "T" subscript on the plim of $\hat{\pi}_T$ is generally required because $\hat{\pi}_T$ need not be estimating any "true" parameters (parameters of interest) even when (2.5) holds. For example, $\hat{\pi}_T$ could be an estimator from a misspecified parameterized conditional variance equation in the context of weighted nonlinear least squares, or it could be an estimator from an alternative model in a nonnested hypotheses framework. When the data are strictly stationary, π_T^0 does not depend on T. In Section 5, when second moment

specification is considered, the plim of $\hat{\pi}_T$ is π_0 where (θ_0, π_0) indexes the conditional variance of Y_t given X_t under the null hypothesis.

The QMLE $\hat{\theta}_T$ maximizes the quasi log-likelihood function

$$L_T(\theta; Y, Z) = \sum_{t=1}^T \ell_t(Y_t, X_t; \hat{\pi}_T, \theta)$$

where $\ell_t(y_t, x_t; \pi, \theta)$ is the conditional log likelihood

$$a(m_t(x_t, \theta), \eta_t(x_t, \pi)) + y_t c(m_t(x_t, \theta), \eta_t(x_t, \pi))$$

(note that $b(y_t, \eta_t(x_t, \hat{\pi}_T))$ does not appear since it does not affect the optimization problem). For simplicity, the argument x_t is omitted wherever convenient. Letting $s_{Tt}(\theta) \equiv \nabla_{\theta} \ell_t(\theta, \hat{\pi}_T^0)$ and suppressing the dependence of s_{Tt} on $\hat{\pi}_T^0$ (and hence on T) we have

$$\begin{aligned} (2.6) \quad s_t(\theta) &= \nabla_m a(m_t(\theta)) \nabla_{\theta} m_t(\theta) + Y_t \nabla_m c(m_t(\theta)) \nabla_{\theta} m_t(\theta) \\ &= (Y_t - m_t(\theta)) \nabla_m c(m_t(\theta)) \nabla_{\theta} m_t(\theta) \\ &\equiv U_t(\theta) \nabla_m c(m_t(\theta)) \nabla_{\theta} m_t(\theta) \end{aligned}$$

where $U_t(\theta) \equiv Y_t - m_t(X_t, \theta)$ is the $1 \times K$ residual function. The second equality follows from (2.2). If the mean is correctly specified then $E(Y_t | X_t) = m_t(X_t, \theta_0)$ and the true residuals $U_t^0 \equiv U_t(\theta_0)$ are defined. In this case, because $\nabla_{\theta} m_t(\theta_0)$ and $\nabla_m c_t(m_t(\theta_0))$ depend only on X_t ,

$$(2.7) \quad E(s_t(\theta_0) | X_t) = 0.$$

This shows that $\{s_t^0 \equiv s_t(\theta_0) : t=1, 2, \dots\}$ is a martingale difference sequence with respect to the σ -fields $\{\sigma(Y_t, X_t) : t=1, 2, \dots\}$.

Equation (2.7) also establishes Fisher-consistency of the QMLE (when $\hat{\pi}_T$ is replaced by its plim, or any fixed value) and is the basis for

the GMT (1984a) results for dynamic models. It can also be shown (see GMT (1984a)) that if the conditional mean is correctly specified then

$$(2.8) \quad E(h_t(\theta_0) | X_t) = -\nabla m_t^0, \nabla c_t^0 \nabla m_t^0$$

where $h_t(\theta) \equiv \nabla_{\theta} s_t(\theta)$ and values with a "o" superscript are evaluated at (θ_0, π_T^0) . The conditional variance of the score is

$$(2.9) \quad V(s_t^0 | X_t) = \nabla m_t^0, \nabla c_t^0 \Omega_t^0(X_t) \nabla c_t^0 \nabla m_t^0$$

where $\Omega_t^0(X_t) \equiv V(Y_t | X_t = x_t)$ is the true conditional covariance matrix of Y_t given X_t . It can be shown that $[\nabla_m c(m_t(\theta), \eta_t(\pi))]^{-1}$ is the covariance matrix associated with the density $f_t(y_t | x_t, m_t(\theta), \eta_t(\pi))$. The conditional information equality holds provided

$$(2.10) \quad \Omega_t^0(X_t) = [\nabla_m c(m_t(X_t, \theta_0), \eta_t(X_t, \pi_T^0))]^{-1} \quad t=1, 2, \dots$$

and this is the case if the assumed density (evaluated at (θ_0, π_T^0)) has second moment corresponding to the actual conditional covariance of Y_t . In general,

$$A_T^0 \equiv -T^{-1} \sum_{t=1}^T E[h_t(\theta_0, \pi_T^0)] \quad \text{and}$$

$$B_T^0 \equiv T^{-1} \sum_{t=1}^T E[s_t(\theta_0, \pi_T^0)' s_t(\theta_0, \pi_T^0)]$$

differ even when the conditional mean is correctly specified, so that the information equality fails. Under standard conditions, $T^{1/2}(\hat{\theta}_T - \theta_0)$ converges in law to $N(0, A_T^{0-1} B_T^0 A_T^{0-1})$. Because A_T^0 and B_T^0 can be consistently estimated by positive semi-definite matrices which require only first derivatives of m_t (with respect to θ) and c (with respect to m), robust classical inference is fairly straightforward for this class of QMLE's. The next section derives

regression-based specification tests which allow robust inference for a wide variety of testing procedures.

3. CONDITIONAL MEAN TESTS

This section focuses on a class of specification tests designed to detect departures from the hypothesis

$$(3.1) \quad H_0: E(Y_t | X_t) = m_t(X_t, \theta_0), \text{ for some } \theta_0 \in \Theta, t=1, 2, \dots$$

Let $U_t(\theta) \equiv Y_t - m_t(\theta)$ be the $1 \times K$ residual function, and let $U_t^0 = U_t(\theta_0)$ be the "true" residuals under H_0 . Suppose that $\Lambda_t(X_t)$ is a $K \times Q$ matrix function of the predetermined variables X_t . If (3.1) holds then by the law of iterated expectations (assuming existence of the appropriate moments),

$$E[U_t^0 \Lambda_t(X_t, \theta_0, \pi_T^0) \nabla c_t(\theta_0, \pi_T^0)] = 0, \quad t=1, 2, \dots$$

Note that Λ_t is allowed to depend on θ_0 and the limiting value of the nuisance parameter estimator.

As pointed out by Newey (1985) and Tauchen (1985), (3.2) suggests basing a test of (3.1) on a quadratic form in the $Q \times 1$ vector

$$\hat{\Psi}_T \equiv T^{-1} \sum_{t=1}^T \psi_t(\hat{\theta}_T, \hat{\pi}_T)$$

where $\psi_t(\theta, \pi) \equiv \Lambda_t(\theta, \pi)' \nabla c_t(\theta, \pi) U_t(\theta)'$. It is readily seen that the asymptotic distribution of $T^{1/2} \hat{\Psi}_T$ does not depend on that of $\hat{\pi}_T$ provided $T^{1/2}(\hat{\pi}_T - \pi_T^0)$ is $O_p(1)$, which is typically the case. Under suitable regularity conditions, it is straightforward to establish that under H_0 ,

$$(3.2) \quad T^{1/2} \hat{\Psi}_T \stackrel{d}{\rightarrow} N(0, \Gamma_T^0 V_T^0 \Gamma_T^{0'})$$

where

$$(3.3) \quad \Gamma_T^0 \equiv T^{-1} \sum_{t=1}^T [I_Q \quad -E(\Lambda_t^0, \nabla c_t^0 \nabla m_t^0) \{E(\nabla m_t^0, \nabla c_t^0 \nabla m_t^0)\}^{-1}]$$

and

$$(3.4) \quad V_T^0 \equiv T^{-1} \sum_{t=1}^T V([\psi_t^0, s_t^0]') = T^{-1} \sum_{t=1}^T E([\psi_t^0, s_t^0]' [\psi_t^0, s_t^0]).$$

Equation (3.2) can be used as the basis for testing the correct specification of the conditional mean, with the resulting statistics being robust against misspecification of other aspects of the conditional distribution of Y_t given X_t . All that one needs are consistent estimators of Γ_T^0 and V_T^0 , and these are available from (3.3), (3.4), \hat{U}_t , $\hat{\Theta}_T$ and $\hat{\pi}_T$. However, computation of the resulting test statistic requires special programming. A method for computing the test statistics that requires only auxiliary least squares regressions can substantially reduce the computational burden and give insight about the directions of misspecification for which the test is inconsistent.

To motivate the general approach, consider a familiar example. For simplicity, assume that the observations are independent so that $X_t = Z_t$, and consider the linear model

$$(3.5) \quad E(Y_t | X_t) = X_{t1} \alpha_0 + X_{t2} \beta_0$$

where Y_t is a scalar, X_{t1} is $1 \times P_1$ and X_{t2} is $1 \times P_2$. Suppose that the hypothesis of interest is

$$(3.6) \quad H_0: \beta_0 = 0.$$

The LM approach leads to a test based on the sample covariance of

the residuals estimated under the null and the excluded variables X_{t2} , i.e.,

$$(3.7) \quad T^{-1} \sum_{t=1}^T X'_{t2} \hat{U}_t.$$

Suppose that, instead of directly using (3.7), the part of X_{t1} that is correlated with X_{t2} is first removed from X_{t2} . That is, perform a multivariate regression of X_{t2} on X_{t1} and form the residuals $\ddot{X}_{t2} \equiv X_{t2} - X_{t1} \hat{B}_{T1}$, $t=1, \dots, T$ where \hat{B}_{T1} is the $P_1 \times P_2$ matrix of least squares coefficients:

$$\hat{B}_{T1} \equiv \left(\sum_{t=1}^T X'_{t1} X_{t1} \right)^{-1} \sum_{t=1}^T X'_{t1} X_{t2}.$$

Then, because $\sum_{t=1}^T X'_{t1} \hat{U}_t = 0$, the statistic in (3.7) is identical to that obtained by replacing X_{t2} with \ddot{X}_{t2} :

$$(3.8) \quad \hat{\vartheta}_T \equiv T^{-1/2} \sum_{t=1}^T \ddot{X}'_{t2} \hat{U}_t = T^{-1/2} \sum_{t=1}^T X'_{t2} \hat{U}_t.$$

The advantage of working with $\hat{\vartheta}_T$ expressed in terms of $\ddot{X}'_{t2} \hat{U}_t$ is that it can be expanded as

$$(3.9) \quad \begin{aligned} \hat{\vartheta}_T &= T^{-1/2} \sum_{t=1}^T (X_{t2} - X_{t1} B_{T1}^0)' U_t^0 \\ &+ T^{-1} \sum_{t=1}^T (X_{t2} - X_{t1} B_{T1}^0)' X_{t1} T^{1/2} (\hat{\alpha}_T - \alpha_0) \\ &+ T^{-1/2} (B_{T1}^0 - \hat{B}_{T1})' \sum_{t=1}^T X'_{t1} \hat{U}_t \end{aligned}$$

where

$$B_{T1}^0 \equiv \left(T^{-1} \sum_{t=1}^T E(X'_{t1} X_{t1}) \right)^{-1} T^{-1} \sum_{t=1}^T E(X'_{t1} X_{t2}).$$

By the first order condition for the OLS estimator $\hat{\alpha}_T$, the third term on the right hand side of (3.9) is identically zero. Also note that B_{T1}^0 is defined so that

$$T^{-1} \sum_{t=1}^T E[(X_{t2} - X_{t1} B_{T1}^0)' X_{t1}] = 0.$$

By the weak law of large numbers (WLLN),

$$T^{-1} \sum_{t=1}^T (X_{t2} - X_{t1} B_{T1}^0)' X_{t1} = o_p(1).$$

Combined with $T^{1/2}(\hat{\alpha}_T - \alpha_0) = O_p(1)$, this shows that the second term in (3.9) is $o_p(1)$ under general conditions if H_0 is true. Deriving the limiting distribution of the LM statistic therefore reduces to deriving the limiting distribution of

$$(3.10) \quad T^{-1/2} \sum_{t=1}^T (X_{t2} - X_{t1} B_{T1}^0)' U_t^0.$$

Under standard regularity conditions, if H_0 is true, (3.10) is asymptotically $N(0, \Xi_T^0)$ where

$$(3.11) \quad \Xi_T^0 = T^{-1} \sum_{t=1}^T E[(U_t^0)^2 (X_{t2} - X_{t1} B_{T1}^0)' (X_{t2} - X_{t1} B_{T1}^0)].$$

Note that Ξ_T^0 is the correct expression whether or not $E[(U_t^0)^2 | X_t]$ is constant. Following White (1980a), a consistent estimator of Ξ_T^0 is

$$(3.13) \quad \hat{\Xi}_T = T^{-1} \sum_{t=1}^T \hat{U}_t^2 (X_{t2} - X_{t1} \hat{B}_{T1})' (X_{t2} - X_{t1} \hat{B}_{T1}).$$

For testing H_0 , this modified LM approach leads to the statistic

$$(3.14) \quad \hat{S}_T' \hat{\Xi}_T^{-1} \hat{S}_T$$

which is distributed asymptotically as χ_{P-2}^2 under H_0 in the presence of heteroskedasticity of unknown form. From a computational viewpoint, it is useful to note that (3.14) is computable as TR^2 , where the R^2 is the uncentered r -squared from the auxiliary regression

$$(3.15) \quad 1 \quad \text{on} \quad \hat{X}_{t2} \hat{U}_t \quad t=1, \dots, T.$$

Interestingly, the regression in (3.15) yields a test statistic that is numerically equivalent to what would be obtained by applying White's (1984, Theorem 4.32) robust form of the LM statistic to the case of heteroskedasticity. The preceding analysis shows that the White statistic can be computed entirely from least squares regressions.

The procedure outlined above can be compared to the standard LM approach. If the assumption of conditional homoskedasticity is maintained and Ξ_T^0 is estimated by

$$(3.16) \quad \hat{\Xi}_T \equiv \hat{\sigma}_T^2 T^{-1} \sum_{t=1}^T (X_{t2} - X_{t1} \hat{B}_{T1})' (X_{t2} - X_{t1} \hat{B}_{T1}),$$

where $\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^T \hat{U}_t^2$, then the resulting test statistic $\hat{\theta}'_{T1} \hat{\Xi}_T^{-1} \hat{\theta}_T$ is

exactly the r-squared form of the LM statistic, which is obtained from the regression

$$(3.17) \quad \hat{U}_t \quad \text{on} \quad X_{t1}, X_{t2} \quad t=1, \dots, T.$$

This regression has an uncentered r-squared which is identical to the uncentered r-squared from the regression

$$(3.18) \quad \hat{U}_t \quad \text{on} \quad X_{t2} - X_{t1} \hat{B}_{T1} \quad t=1, \dots, T$$

verifying that the robust statistic obtained from the regression in (3.15) is asymptotically equivalent to the statistic obtained from (3.17) (or (3.18)) if H_0 is true and Y_t is conditionally homoskedastic. In general, as emphasized by White (1980a,b,1984) in several contexts, the regression statistic based on (3.17) is not

asymptotically χ_{P-2}^2 under H_0 if heteroskedasticity is present. The statistic based on the regression in (3.15) does have a limiting χ_{P-2}^2 under H_0 in the presence of heteroskedasticity of unknown form. Thus the statistic based on (3.15) is preferred.

To extend the above approach for general CM tests, recall the first order condition for the QMLE $\hat{\theta}_T$:

$$(3.20) \quad \sum_{t=1}^T \nabla_{\theta} m_t(\hat{\theta}_T)' \nabla_m c_t(m_t(\hat{\theta}_T), \eta_t(\hat{\pi}_T)) (Y_t - m_t(\hat{\theta}_T))' \equiv 0$$

or, in shorthand,

$$(3.21) \quad \sum_{t=1}^T \nabla m_t' \nabla c_t \hat{U}_t' = 0.$$

Therefore, the indicator $\hat{\Psi}_T \equiv T^{-1} \sum_{t=1}^T \hat{\Lambda}_t' \nabla c_t \hat{U}_t'$ is identical to

$$(3.22) \quad T^{-1} \sum_{t=1}^T (\nabla c_t^{1/2} \hat{\Lambda}_t - \nabla c_t^{1/2} \nabla m_t \hat{B}_T)' \nabla c_t^{1/2} \hat{U}_t'$$

where \hat{B}_T is the $P \times Q$ matrix of regression coefficients from a matrix regression of $\nabla c_t^{1/2} \hat{\Lambda}_t$ on $\nabla c_t^{1/2} \nabla m_t$:

$$(3.23) \quad \hat{B}_T = \left(\sum_{t=1}^T \nabla m_t' \nabla c_t \nabla m_t \right)^{-1} \sum_{t=1}^T \nabla m_t' \nabla c_t \hat{\Lambda}_t.$$

The statistic in (3.22) first purges from $\nabla c_t^{1/2} \hat{\Lambda}_t$ its least squares projection based on $\nabla c_t^{1/2} \nabla m_t$ before constructing the indicator.

Note that $\nabla c_t^{1/2}$ is an estimator of $[V(Y_t | X_t)]^{-1/2}$ if the second moment is correctly specified, but not in general. It can be shown as in the linear least squares case that

$$(3.24) \quad \begin{aligned} & T^{-1/2} \sum_{t=1}^T (\nabla c_t^{1/2} \hat{\Lambda}_t - \nabla c_t^{1/2} \nabla m_t \hat{B}_T)' \nabla c_t^{1/2} \hat{U}_t' \\ &= T^{-1/2} \sum_{t=1}^T (\nabla c_{ot}^{1/2} \Lambda_t^o - \nabla c_{ot}^{1/2} \nabla m_t^o B_T^o)' \nabla c_{ot}^{1/2} U_t^o + o_p(1), \end{aligned}$$

where

$$B_T^0 \equiv \left(T^{-1} \sum_{t=1}^T E(\nabla m_t^0, \nabla c_t^0 \nabla m_t^0) \right)^{-1} T^{-1} \sum_{t=1}^T E(\nabla m_t^0, \nabla c_t^0 \Lambda_t^0)$$

and values with "o" superscripts are evaluated at (θ_0, π_T^0) . Under H_0 , a consistent estimator of the asymptotic covariance matrix of the right hand side of (3.24) (and therefore of the LHS as well) is

$$(3.25) \quad T^{-1} \sum_{t=1}^T (\nabla c_t^{1/2} \hat{\Lambda}_t - \nabla c_t^{1/2} \hat{\nabla} m_t \hat{B}_T), \nabla c_t^{1/2} \hat{U}_t \hat{U}_t' \nabla c_t^{1/2} (\nabla c_t^{1/2} \hat{\Lambda}_t - \nabla c_t^{1/2} \hat{\nabla} m_t \hat{B}_T).$$

Equations (3.24) and (3.25) lead to the following theorem.

Theorem 3.1: Suppose that the following assumptions hold:

- (i) Regularity conditions A.1 in the appendix.
- (ii) $H_0: E(Y_t | X_t) = m_t(X_t, \theta_0)$, for some $\theta_0 \in \Theta$, $t=1, 2, \dots$

Then

$$\xi_T \stackrel{d}{\rightarrow} \chi^2_{\bar{r}}$$

where $\xi_T = TR^2$ and R^2 is the uncentered r-squared from the auxiliary regression

$$1 \quad \text{on} \quad \hat{U}_t \nabla c_t (\hat{\Lambda}_t - \hat{\nabla} m_t \hat{B}_T) \quad t=1, \dots, T,$$

and \hat{B}_T is given by (3.23). ■

In practice, Theorem 3.1 can be applied as follows:

- (i) Given the nuisance parameter estimate $\hat{\pi}_T$, compute the QMLE $\hat{\theta}_T$, the weighted residuals $\tilde{U}_t \equiv \nabla c_t^{1/2} \hat{U}_t$, the weighted regression function $\tilde{\nabla} m_t \equiv \nabla c_t^{1/2} \hat{\nabla} m_t$, and the weighted indicator function $\tilde{\Lambda}_t \equiv \nabla c_t^{1/2} \hat{\Lambda}_t$;

(ii) Perform a multivariate regression of $\tilde{\Lambda}_t$ on $\tilde{\nabla m}_t$ and keep the residuals, say $\tilde{\Lambda}_t$:

(iii) Perform the OLS regression

$$(3.26) \quad 1 \quad \text{on} \quad \tilde{U}_t \tilde{\Lambda}_t \quad t=1,2,\dots,T$$

and use TR^2 as asymptotically χ^2_Q under H_0 , where R^2 is of course the uncentered r-squared. ■

This procedure assumes that the matrix

$$T^{-1} \sum_{t=1}^T E[(\Lambda_t^0 - \nabla_{\theta_t^0} B_T^0)' \nabla_{m_t^0} U_t^0 U_t^0 \nabla_{m_t^0} (\Lambda_t^0 - \nabla_{\theta_t^0} B_T^0)']$$

is positive definite uniformly in T . This condition can fail if the weighted indicator $\nabla_{m_t^0}^{01/2} \Lambda_t^0$ contains redundancies (more precisely, if $\nabla_{m_t^0}^{01/2} \Lambda_t^0 - \nabla_{m_t^0}^{01/2} \nabla_{\theta_t^0} B_T^0$ contains redundancies). In this case the regression form can still be used, but the degrees of freedom in the chi-square distribution must be appropriately reduced.

The above procedure gives a simple method for testing specification of the conditional mean for a broad class of multivariate models estimated by QMLE, without imposing the additional assumption that the conditional variance of Y_t is correctly specified under the null hypothesis. Note that if the second moment of Y_t is correctly specified then the weighting matrix appearing above is the negative square root of the conditional variance of Y_t . In general, the weight corresponds to the variance of the assumed distribution, and need not equal the conditional covariance matrix of Y_t given X_t .

The regression appearing in (3.26) is similar to auxiliary regressions appearing in Newey (1985), White (1985b), and elsewhere,

but there is an important difference. Consider the regressions

$$(3.27) \quad 1 \quad \text{on} \quad \hat{U}_t \nabla c_t \nabla m_t, \hat{U}_t \nabla c_t \hat{\Lambda}_t$$

$$(3.28) \quad \nabla c_t^{1/2} \hat{U}_t \quad \text{on} \quad \nabla c_t^{1/2} \nabla m_t, \nabla c_t^{1/2} \hat{\Lambda}_t$$

where the multivariate regression in (3.28) (when $K > 1$) is carried out by stacking the data and using OLS. If $[\nabla c_t(\theta_0, \pi_T^0)]^{-1} = V(Y_t | X_t)$ then the TR^2 from either of these regressions is asymptotically distributed as χ_Q^2 under H_0 . If the conditional second moment of Y_t is misspecified, i.e. $[\nabla c_t(\theta_0, \pi_T^0)]^{-1} \neq V(Y_t | X_t)$, then neither of these statistics generally has an asymptotic χ_Q^2 distribution (although the statistic obtained from (3.27) has a better chance of having a limiting chi-square distribution), and they are no longer asymptotically equivalent. The statistic derived from (3.26) has a limiting χ_Q^2 distribution under H_0 whether or not the second moment is correctly specified. If the conditional mean is the object of interest, and the researcher is at all uncertain about the distributional assumption, then the methodology of Theorem 3.1 is preferred.

The setup of Theorem 3.1 allows consideration of a wide class of procedures used by applied econometricians. Some examples of how to choose Λ_t in some familiar cases are given in the following section.

4. EXAMPLES OF CONDITIONAL MEAN TESTS

The approach taken in Theorem 3.1 was motivated by considering the LM test for exclusion restrictions in a linear model with independent observations. This section presents some further applications of Theorem 3.1. Included are robust LM tests, robust regression forms of Hausman tests, and a modified Davidson-MacKinnon test of nonnested hypotheses.

Example 4.1 (LM test in nonlinear regression): For simplicity, assume that $K=1$, so that Y_t is a scalar. Consider estimating $E(Y_t|X_t)$ by nonlinear least squares. In this case, the nuisance parameter $\pi = \sigma^2$ is the variance associated with the assumed LEF distribution, $N(\mu_t(X_t, \alpha, \beta), \sigma^2)$. Here α is $P_1 \times 1$, β is $P_2 \times 1$. Assume that $E(Y_t|X_t) = \mu_t(X_t, \alpha_0, \beta_0)$. The null hypothesis is

$$H_0: \beta_0 = 0.$$

The LM approach leads to a statistic based upon the $P_2 \times 1$ vector

$$\sum_{t=1}^T \nabla_{\beta} \mu_t(X_t, \hat{\alpha}_T, 0)' \hat{U}_t$$

where $\hat{\alpha}_T$ is the NLLS estimator of α_0 obtained under the assumption that $\beta_0 = 0$, and $\hat{U}_t = Y_t - \mu_t(X_t, \hat{\alpha}_T, 0)$. Let $\nabla_{\alpha} \hat{\mu}_t$, $\nabla_{\beta} \hat{\mu}_t$ denote the gradients of μ_t with respect to α , β evaluated at $(\hat{\alpha}_T', 0)'$. In the notation of Theorem 3.1, $\Theta \equiv \alpha$, $m_t(\Theta) \equiv \mu_t(\alpha, 0)$, $c(m, \eta) = m/\eta$ and $\eta(\pi) \equiv \pi \equiv \sigma^2$. Also, the indicator $\Lambda_t(\alpha, \pi)$ is $\nabla_{\beta} \mu_t(\alpha, 0)$ (where X_t has been suppressed). A test of H_0 which is robust in the presence of heteroskedasticity can be easily computed using the following procedure.

- (i) Estimate α_0 by NLLS assuming $\beta_0 = 0$. Compute the residuals \hat{U}_t , and the gradients $\nabla_{\alpha} \mu_t(\hat{\alpha}_T, 0)$ and $\nabla_{\beta} \mu_t(\hat{\alpha}_T, 0)$.
- (ii) Regress $\nabla_{\beta} \hat{\mu}_t$ on $\nabla_{\alpha} \hat{\mu}_t$ and keep the residuals, say $\nabla_{\beta} \ddot{\mu}_t$.
- (iii) Regress 1 on $\hat{U}_t \nabla_{\beta} \ddot{\mu}_t$ and use TR^2 from this regression as asymptotically $\chi_{F_2}^2$ under H_0 .

As a special case of this procedure, consider the LM test for AR(1) serial correlation in a dynamic model. The null and alternative hypotheses are

$$H_0: E(Y_t | X_t) = m_t(X_t, \alpha_0) \quad t=1, 2, \dots$$

$$H_1: E(Y_t | X_t) = m_t(X_t, \alpha_0) + \rho_0(Y_{t-1} - m_{t-1}(X_{t-1}, \alpha_0)) \quad t=2, 3, \dots$$

The LM procedure leads to a test based upon

$$\sum_{t=2}^T \hat{U}_t \hat{U}_{t-1}$$

where $\hat{U}_t = Y_t - m_t(X_t, \hat{\alpha}_T)$ and $\hat{\alpha}_T$ is the NLLS estimator obtained under the assumption of no serial correlation. Theorem (3.1) leads to the following procedure:

- (i) Estimate α_0 by NLLS. Keep \hat{U}_t , $\nabla \hat{m}_t \equiv \nabla_{\alpha} m_t(X_t, \hat{\alpha}_T)$.
- (ii) Regress \hat{U}_{t-1} on $\nabla \hat{m}_t$ and keep the residuals from this regression, say \ddot{U}_{t-1} .
- (iii) Regress 1 on $\hat{U}_t \ddot{U}_{t-1}$ and use $(T-1)R^2$ from this regression as asymptotically χ_1^2 under H_0 .

Note that this procedure assumes nothing about the conditional variance of Y_t given X_t . Also, X_t may contain lagged values of Y_t , as well as Z_t that are not strictly exogenous. This procedure maintains the spirit of the usual LM procedure, but is robust to

heteroskedasticity. The OLS regression in (ii) is the cost to being robust to heteroskedasticity.

Extending the above procedure to test for AR(Q) serial correlation is straightforward. In (ii), regress $\hat{U}_{t-1}, \dots, \hat{U}_{t-Q}$ on $\nabla \hat{m}_t$ and save the residuals \hat{U}_{t-j} , and in (iii), regress 1 on $\hat{U}_t \hat{U}_{t-j}$, $j=1, \dots, Q$. $(T-Q)R^2$ from this regression is asymptotically χ_Q^2 under H_0 .

The above analysis extends to the case that the restrictions cannot be written as exclusion restrictions. Let δ_0 be the $(P+Q) \times 1$ vector of parameters in the unconstrained model and suppose that the restrictions under H_0 can be expressed as

$$\delta_0 = r(\alpha_0) \quad \text{for some } \alpha_0 \in A$$

where $A \subset \mathbb{R}^F$ and $r: A \rightarrow \Delta$. Let $\mu_t(x_t, \alpha) \equiv m_t(x_t, r(\alpha))$. If α_0 is in the interior of A and r is differentiable on $\text{int } A$ then a heteroskedasticity-robust test of H_0 is obtained as follows:

- (i) Estimate α_0 by NLLS and save $\nabla_{\alpha} \mu_t(\hat{\alpha}_T)$ and the residuals \hat{U}_t . Let $\hat{\delta}_T \equiv r(\hat{\alpha}_T)$ be the constrained estimator of δ_0 ;
- (ii) Let $\nabla_{\delta} \hat{m}_t \equiv \nabla_{\delta} m_t(\hat{\delta}_T)$ and run the multivariate regression

$$\nabla_{\delta} \hat{m}_t \quad \text{on} \quad \nabla_{\alpha} \hat{\mu}_t \quad t=1, \dots, T$$

and save the residuals, $\nabla_{\delta} \hat{m}_t$;

- (iii) Run the regression

$$1 \quad \text{on} \quad \nabla_{\delta} \hat{m}_t \quad t=1, \dots, T$$

and use TR^2 as asymptotically χ_Q^2 under H_0 . ■

Note that there is perfect multicollinearity in the regression in (iii), so that P of the indicators can be excluded if the

regression package used does not compute R^2 's for regressions containing perfect multicollinearity. Also, note that there is no need to explicitly compute the gradient of r with respect to α . This is to be contrasted with other methods to compute heteroskedasticity-robust test statistics (e.g. White (1984, Chapter 4)). ■

Example 4.2 (Hausman test for a conditional mean): Suppose, in the spirit of the Hausman (1978) methodology, two estimators of the conditional mean parameters θ_0 are compared in order to detect misspecification of the regression function. Because the QMLE's considered here yield consistent estimators of the conditional mean, it is natural to base a test on the difference of two such estimators. A regression form of the test can be derived which does not require either estimator to be the efficient QMLE. Also, only one of the QMLE's needs to be computed.

Suppose that $\hat{\theta}_T$ is the QMLE from an LEF indexed by (a_1, b_1, c_1) and nuisance parameters $\eta_1(\pi_1)$ and a second estimator is to be used from the LEF (a_2, b_2, c_2) with nuisance parameters $\eta_2(\pi_2)$. Then a statistic that is asymptotically equivalent to the Hausman test which directly compares the two estimators is obtained by taking

$$(4.1) \quad \Lambda_t(\theta, \pi) \equiv [\nabla_m c_1(m_t(\theta), \eta_{t1}(\pi_1))]^{-1} \nabla_m c_2(m_t(\theta), \eta_{t2}(\pi_2)) \nabla_{\theta} m_t(\theta).$$

In the notation of Theorem 3.1, $\pi \equiv (\pi'_1, \pi'_2)'$ and $c(m, \eta) \equiv c_1(m, \eta_1)$. Let $\hat{\pi}_{T1}$ and $\hat{\pi}_{T2}$ denote the nuisance parameter estimators. The procedure for carrying out the Hausman test is as follows.

(i) Given the nuisance parameter estimate $\hat{\pi}_{T1}$, compute the QMLE $\hat{\Theta}_T$ using the LEF (a_1, b_1, c_1) . From $\hat{\Theta}_T$, $\hat{\pi}_{T1}$ and $\hat{\pi}_{T2}$, compute the weighted residuals, $\tilde{U}_t \equiv \hat{U}_t \hat{\nabla} c_{t1}^{1/2}$, the weighted regression function $\tilde{\nabla} m_t \equiv \hat{\nabla} c_{t1}^{1/2} \hat{\nabla} m_t$, and the weighted indicator function $\tilde{\Lambda}_t \equiv \hat{\nabla} c_{t1}^{1/2} \hat{\Lambda}_t = \hat{\nabla} c_{t1}^{-1/2} \hat{\nabla} c_{t2} \hat{\nabla} m_t$ (note that the indicator is just another weighting of the regression function).

(ii) Perform the multivariate regression of

$$\tilde{\Lambda}_t \quad \text{on} \quad \tilde{\nabla} m_t$$

and keep the residuals, say $\ddot{\Lambda}_t$.

(iii) Perform the regression

$$1 \quad \text{on} \quad \tilde{U}_t \ddot{\Lambda}_t \quad t=1, \dots, T$$

and use TR^2 from this regression as Y_P^2 under H_0 where P is the dimension of Θ_0 .

For concreteness, consider a special case. Suppose Y_t is a scalar count variable, and the researcher postulates the conditional mean function

$$(4.2) \quad H_0: E(Y_t | X_t) = \exp(W_t \theta_0)$$

where W_t is a $1 \times P$ subvector of X_t with a maximum lag length that is independent of t (for cross section applications, W_t is a subset of Z_t). Because Y_t is a count variable, it is sensible to use a Poisson likelihood function to estimate θ_0 . However, for Y_t to truly have a conditional Poisson distribution, its conditional mean and conditional variance must be equal; this is a fairly strong restriction which can yield misleading results if it is violated. For now, assume that interest lies only in testing H_0 . If H_0 is

true, the Poisson QMLE and NLLS both consistently estimate θ_0 ; this is the basis for a Hausman test.

For the Poisson model, there are no nuisance parameters. In the previous notation, $a_1(m) \equiv -m$, $b_1(y) \equiv -\log(y!)$ and $c_1(m) \equiv \log m$. For NLLS, the nuisance parameter is the variance associated with the normal density, so $\eta_2(\pi_2) \equiv \sigma_2^2$, $a_2(m, \sigma_2^2) \equiv -m^2/2\sigma_2^2$, $b_2(y, \sigma_2^2) \equiv -(y^2 + \log(2\pi))/\sigma_2^2$, and $c_2(m, \sigma_2^2) \equiv m/\sigma_2^2$. Then

$$\nabla_{m_t}(\theta) = \exp(W_t\theta)W_t, \quad \nabla_{c_{t1}}(\theta) = \exp(-W_t\theta), \quad \nabla_{c_{t2}}(\theta, \sigma_2^2) = 1/\sigma_2^2$$

and

$$\nabla_{c_{t1}}^{1/2} \hat{\Lambda}_t = \exp(3/2W_t \hat{\theta}_T)W_t / \sigma_2^2, \quad \nabla_{c_{t1}}^{1/2} \nabla_{m_t} \hat{\Lambda}_t = \exp(1/2W_t \hat{\theta}_T)W_t$$

where $\hat{\theta}_T$ is the Poisson QMLE. The estimate of σ_2^2 can be ignored since it appears only as a scaling factor. Define the residuals $\hat{U}_t \equiv Y_t - \exp(W_t \hat{\theta}_T)$ and the weighted residuals $\tilde{U}_t \equiv \exp(-1/2W_t \hat{\theta}_T) \hat{U}_t$. Perform the multivariate regression

$$\exp(3/2W_t \hat{\theta}_T)W_t \quad \text{on} \quad \exp(1/2W_t \hat{\theta}_T)W_t$$

and keep the residuals, say $\ddot{\Lambda}_t$. Then perform the regression

$$1 \quad \text{on} \quad \tilde{U}_t \ddot{\Lambda}_t$$

and use TR^2 as χ_F^2 under H_0 .

Again, it is emphasized that this procedure does not assume that the Poisson distribution (or the normal) is correctly specified under H_0 . This is in contrast to the usual method used to compute the Hausman test in this context. Hausman, Ostro and Wise (1984) (HOW) apply a regression method which is very similar to what would be obtained from (3.28) in the Poisson context. The difference is that they perform the NLLS estimation so that the roles of the two

QMLE's are reversed (it is straightforward to carry out the robust test when θ_0 is estimated by NLLS rather than Poisson QMLE). The assumption that the Poisson distribution is correct leads to a regression test that is slightly different than (3.28); essentially, the procedure is to perform an LM test for exclusion of $\exp(1/2W_t\hat{\theta}_T)W_t$ in the NLLS model (also see White (1985b, pp.25-26)). If the Poisson assumption is incorrect, then this approach leads to a test with incorrect asymptotic size for testing H_0 . In addition, the HOW procedure is not consistent for testing the Poisson distributional specification in the following sense: if the conditional mean is correct but the Poisson assumption is violated, the HOW test statistic still has a well-defined limiting distribution (which is not χ^2_P), rather than tending in probability to infinity. This result extends to all Hausman tests based on two QMLE's that are derived from the LEF class of densities. My opinion is that the Hausman test in the present context should not be viewed as a general test of distributional misspecification. A test which is useful for testing distributional assumptions beyond the first moment is derived in the next section. For the Poisson case, it leads to comparing the estimated conditional mean and another estimate of the conditional variance. ■

Example 4.3 (Robust Davidson-MacKinnon nonnested hypotheses test):

Let Y_t be a random scalar, let X_t be the predetermined variables as defined in Section 2. Consider two competing models for $E(Y_t|X_t)$:

$$H_0: E(Y_t|X_t) = m_t(X_t, \theta_0), \text{ some } \theta_0 \in \Theta, t=1,2,\dots$$

$$H_1: E(Y_t|X_t) = \mu_t(X_t, \delta_0), \text{ some } \delta_0 \in \Delta, t=1,2,\dots$$

The DM test looks for departures from H_0 in the direction of H_1 (of course the roles of H_0 and H_1 can be reversed). Assume that model H_0 is estimated by NLLS. Let $\hat{\theta}_T$ denote the NLLS estimator of θ_0 . When H_0 is true, NLLS on model H_1 will yield an estimator $\hat{\delta}_T$ which generally converges to $\delta_T^0 \in \Delta$ (note that δ_T^0 does not have an interpretation as "true" parameters, but it produces the smallest mean squared error approximation to $E(Y_t|X_t)$ in the parametric class $\{\mu_t(x_t, \delta) : \delta \in \Delta\}$). The estimated "nuisance parameters" in this setup are $\hat{\pi}_T \equiv (\hat{\sigma}_T^2, \hat{\delta}_T)$ where $\hat{\sigma}_T^2$ is the estimated "variance" under H_0 .

The DM test checks for nonzero correlation between the residuals $\hat{U}_t = (Y_t - m_t(X_t, \hat{\theta}_T))$ and the difference in the estimated regression functions, $\mu_t(X_t, \hat{\delta}_T) - m_t(X_t, \hat{\theta}_T)$. In the notation of Theorem 3.2,

$$\nabla c_t(m_t(\theta), \eta_t(\pi)) \equiv 1/\sigma^2$$

and
$$\Lambda_t(\theta, \pi) = \mu_t(\delta) - m_t(\theta).$$

If $V(Y_t|X_t) = \sigma_0^2$ is maintained, then a regression test is available from (3.28): run the regression

$$\hat{U}_t \quad \text{on} \quad \nabla \hat{m}_t, \mu_t(\hat{\delta}_T) - m_t(\hat{\theta}_T)$$

and use TR^2 as asymptotically χ_1^2 under H_0 . This is the LM form of

the usual Davidson-MacKinnon test. If $V(Y_t|X_t)$ is not constant under H_0 , then this approach generally leads to inference with the wrong asymptotic size. The robust approach is

(i) Estimate model H_0 by NLLS; save $m_t(X_t, \hat{\theta}_T)$, $\nabla_{\theta} m_t(X_t, \hat{\theta}_T)$, and the residuals $\hat{U}_t \equiv Y_t - m_t(X_t, \hat{\theta}_T)$. Estimate model H_1 by NLLS and save $\mu_t(X_t, \hat{\delta}_T)$.

(ii) Regress $\mu_t(\hat{\delta}_T) - m_t(\hat{\theta}_T)$ on $\nabla m_t(\hat{\theta}_T)$, $t=1, 2, \dots, T$ and save the residuals, say $\hat{\Lambda}_t$.

(iii) Regress 1 on $\hat{U}_t \hat{\Lambda}_t$ and use TR^2 as χ_1^2 under H_0 .

This approach yields correct asymptotic inference under H_0 in the presence of arbitrary forms of heteroskedasticity, and requires only one additional OLS regression. The OLS regression in step (ii) is the cost of the robust procedure. This heteroskedasticity-robust version should be useful in a variety of economic contexts, particularly when the dependent variable is restricted to be nonnegative. In such cases, homoskedasticity is usually an implausible assumption. Rather than compare two separate functional forms for the dependent variable (which, to perform the test correctly, requires a distributional assumption), one can compute two NLLS estimators using the same dependent variable and use the heteroskedasticity robust form. Regression-based versions of the DM test for multivariate models estimated by nonlinear SUR which do not assume $V(Y_t|X_t)$ is constant are also available from Theorem 3.1.

5. TESTING SECOND MOMENT ASSUMPTIONS

The tests developed in Sections 3 and 4 are explicitly designed to detect misspecification of the conditional mean without making the additional assumption that the distribution used in estimation is correctly specified. In Section 3, two regression forms of the conditional moment tests were presented (see (3.27) and (3.28)) that could be guaranteed to have a limiting χ^2 distribution only if the second moment of the assumed distribution matches the true conditional second moment of Y_t . More precisely, the regression tests in (3.27) and (3.28) effectively take the null hypothesis to be

$$H'_0: E(Y_t|X_t) = m_t(X_t, \theta_0) \text{ and } V(Y_t|X_t) = [\nabla_{c_t}(m_t(\theta_0), \eta_t(\pi_0))]^{-1}$$

for some $\theta_0 \in \Theta$, some $\pi_0 \in \Pi$.

The nuisance parameters π_0 are no longer indexed by T because they are now "true" parameters which, along with θ_0 , index the conditional second moment of Y_t . For the validity of the regression tests in (3.27) and (3.28), the correct specification of the second moment is usually needed to consistently estimate the covariance matrix which appears in the conditional moment test statistics. Because violation of the distributional assumption does not lead to inconsistent estimates of a correctly specified mean, the CM tests for the conditional mean based on (3.27) or (3.28) are inconsistent for the alternative

$$H'_1: E(Y_t | X_t) = m_t(X_t, \theta_0) \text{ for some } \theta_0 \in \Theta \text{ but } V(Y_t | X_t) \neq [\nabla c_t(m_t(\theta_0), \eta_t(\pi))]^{-1} \text{ for all } \pi \in \Pi.$$

Because the Hausman test is a CM test, it is also inconsistent for testing distributional aspects beyond the first moment. A more powerful test is needed for detecting departures from the second moment assumption.

Such a test is obtained by applying White's (1982) information matrix (IM) testing principle. Actually, the focus here is on second moments, so that the test derived below is closer to the White (1980a) test for heteroskedasticity extended to nonlinear regression models. The difference is that the tests derived below are computable from linear regressions while taking only H'_0 as the null hypothesis (i.e. there is no need to add auxiliary assumptions under the null in order to obtain a simple regression form of the test).

In the spirit of White (1980a), a test is based on two consistent estimators of the asymptotic covariance matrix of $\hat{\theta}_T$ under the hypothesis that the first two conditional moments are correctly specified. Thus, interest lies in second moment misspecification which invalidates the use of the usual standard errors calculated for the QMLE $\hat{\theta}_T$ (although the approach easily generalizes to more general second moment tests). Let $\hat{\pi}_T$ be the nuisance parameter estimator, and let all other quantities be defined as before.

Consider the difference

$$(5.1) \quad T^{-1} \sum_{t=1}^T \widehat{\nabla m}'_t \widehat{\nabla c}_t \widehat{U}'_t \widehat{U}_t \widehat{\nabla c}_t \widehat{\nabla m}_t - T^{-1} \sum_{t=1}^T \widehat{\nabla m}'_t \widehat{\nabla c}_t \widehat{\nabla m}_t$$

where all quantities are evaluated at $(\widehat{\theta}_T, \widehat{\pi}_T)$ (recall that \widehat{U}_t and $\widehat{\nabla m}_t$ do not depend on $\widehat{\pi}_T$). If the first two conditional moments of Y_t given X_t are correctly specified, then the difference in (5.1) tends to zero in probability as $T \rightarrow \infty$ and the standard errors for $\widehat{\theta}_T$ which are computed under the information matrix equality will be asymptotically valid. If either the conditional mean or variance is misspecified then (5.1) will typically not converge to zero (although there are directions of misspecification for which this statistic will still tend to zero). Thus, a test of correctness of the first two moments can be based on a suitably standardized version of (5.1). Taking the vec of the t th difference in (5.1) yields

$$(5.2) \quad \text{vec}(\widehat{\nabla m}'_t \widehat{\nabla c}_t \widehat{U}'_t \widehat{U}_t \widehat{\nabla c}_t \widehat{\nabla m}_t - \widehat{\nabla m}'_t \widehat{\nabla c}_t \widehat{\nabla m}_t)' = \\ \text{vec}(\widehat{U}'_t \widehat{U}_t - \widehat{\nabla c}_t^{-1})' [(\widehat{\nabla c}_t \widehat{\nabla m}_t) \otimes (\widehat{\nabla c}_t \widehat{\nabla m}_t)]$$

where the relationship $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$ for conformable matrices has been used. Under H'_0 , $E(U_t^D, U_t^D | X_t) = [\nabla c_t(\theta_0, \pi_0)]^{-1}$.

In general, a statistic based on (5.2) will have a limiting distribution that depends on the limiting distribution of $(\widehat{\theta}_T, \widehat{\pi}_T)$. Because $\widehat{\pi}_T$ may come from a variety of sources, this dependence makes general derivation of the limiting distribution of the IM statistic tedious. More importantly, the resulting test statistic is computationally burdensome. A statistic which does not depend on

the limiting distribution of $(\hat{\Theta}_T, \hat{\pi}_T)$ would be particularly convenient in this case.

To derive such a statistic, the approach used in Section 3 is modified. Note that $[\nabla c_t(\Theta_0, \pi_0)]^{-1}$ plays the same role for U_t^0, U_t^0 that $m_t(\Theta_0)$ plays for Y_t . Define

$$x_t(\Theta, \pi) \equiv \text{vec} [\nabla_m c_t(m_t(\Theta), \eta_t(\pi))]^{-1},$$

$$\Lambda_t(\Theta, \pi) \equiv [(\nabla c_t(\Theta, \pi) \nabla m_t(\Theta, \pi)) \otimes (\nabla c_t(\Theta, \pi) \nabla m_t(\Theta, \pi))].$$

In Section 3, a statistic which did not depend on the limiting distribution of the parameter estimates was obtained by first removing the influence of the gradient of the conditional mean on the indicator. Things are more complicated here because U_t^0, U_t^0 is no longer observed, but only estimated. Nevertheless, because the gradient of $U_t(\Theta)' U_t(\Theta)$ evaluated at Θ_0 is uncorrelated with any function of X_t , the same strategy works.

If $x_t(\Theta, \pi)$ has a zero derivative with respect to a certain parameter then the estimator of this parameter has no effect asymptotically on the usual IM statistic. In what follows, $\nabla x_t(\Theta, \pi)$ contains only the nonzero, nonredundant elements of the gradient of x_t with respect to both Θ and π . Define

$$\hat{\Gamma}_T \equiv \left(\sum_{t=1}^T \nabla \hat{x}_t' \nabla \hat{x}_t \right)^{-1} \sum_{t=1}^T \nabla \hat{x}_t' \hat{\Lambda}_t,$$

and let $\varphi_t(\Theta, \pi, \Gamma)$ denote the Q nonredundant elements of

$$\text{vec}[U_t(\Theta)' U_t(\Theta) - x_t(\Theta, \pi)]' [\Lambda_t(\Theta, \pi) - \nabla x_t(\Theta, \pi) \Gamma].$$

Theorem 5.1: Suppose that the following conditions hold:

(i) Regularity conditions B.1 in the appendix;

(ii) For some $\theta_0 \in \Theta$, $\pi_0 \in \Pi$, and $t=1,2,\dots$,

$$E(Y_t | X_t) = m_t(X_t, \theta_0)$$

and

$$V(Y_t | X_t) = [\nabla_{\pi} c_t(m_t(X_t, \theta_0), \eta_t(X_t, \pi_0))]^{-1}.$$

Then

$$\xi_T \xrightarrow{d} \chi_Q^2$$

where ξ_T is TR^2 from the regression

$$1 \quad \text{on} \quad \hat{\varphi}_t, \quad t=1,\dots,T$$

and $\hat{\varphi}_t \equiv \varphi_t(\hat{\theta}_T, \hat{\pi}_T, \hat{\Gamma}_T)$. ■

The method for applying Theorem 5.1 is as follows:

(i) Given $\hat{\pi}_T$, compute the QMLE $\hat{\theta}_T$, \hat{U}_t , $\nabla \hat{y}_t$, and $\hat{\Lambda}_t$ as defined above;

(ii) Perform the multivariate regression of $\hat{\Lambda}_t$ on $\nabla \hat{y}_t$ and save the residuals, say $\hat{\Lambda}_t$;

(iii) Perform the OLS regression

$$1 \quad \text{on} \quad \hat{\varphi}_t \quad t=1,\dots,T$$

and use TR^2 as χ_Q^2 , where $\hat{\varphi}_t$ is $1 \times Q$ and contains nonredundant elements of $\text{vec}(\hat{U}_t' \hat{U}_t - \nabla c_t^{-1})' \hat{\Lambda}_t$. ■

Theorem 5.1 gives a simple method for testing correct specification of the conditional second moment. Only least squares regressions are needed to compute the statistic. Calculation of $\nabla \hat{y}_t$ is typically straightforward. The parameters π_0 need not be

estimated by an efficient procedure, so that Theorem 5.1 is applicable to the negative binomial generalization of the Poisson regression model as developed by GMT (1984b). As is the case with the usual IM test, one may exclude indicators and appropriately reduce the degrees of freedom in the χ^2 distribution.

Example 5.1: It is interesting to apply Theorem 5.1 to a univariate nonlinear model that has been estimated by NLLS. In this case, $y_t(\theta, \pi) \equiv \pi \equiv \sigma^2$ so that the only nonzero element of ∇y_t is 1. The indicator $\hat{\Lambda}_t$ is $[\nabla_{m_t}(X_t, \hat{\theta}_T)' \nabla_{m_t}(X_t, \hat{\theta}_T)] / \hat{\sigma}_T^2$ where $\hat{\sigma}_T^2$ is the estimator $T^{-1} \sum_{t=1}^T \hat{U}_t^2$. Theorem 5.1 leads to the regression

$$(5.1) \quad 1 \text{ on } (\hat{U}_t^2 - \hat{\sigma}_T^2)(\hat{\zeta}_t - \bar{\zeta}_T)$$

where $\hat{\zeta}_t$ is a vector of nonconstant, nonredundant elements of $\nabla_{m_t}' \nabla_{m_t}$ and $\bar{\zeta}_T \equiv T^{-1} \sum_{t=1}^T \hat{\zeta}_t$. This procedure is asymptotically equivalent to

the regression form of the White test for heteroskedasticity for nonlinear regression models under the additional assumption that $E[(U_t^0)^4 | X_t]$ is constant (see Domowitz and White (1982)).

Interestingly, the slight modification in (5.1) (which is essentially the demeaning of the indicators $\hat{\zeta}_t$) yields an asymptotically χ^2 distributed statistic without the additional assumption of constant fourth moment for U_t^0 . In the case of a linear time series model, the demeaning of the indicators yields a statistic which is asymptotically equivalent to Hsieh's (1983) robust form of the White test, but the above statistic is significantly easier to compute. Rarely would we care to assume anything about the fourth moment of Y_t , so that the robust

regression form in (5.1) seems to be a useful modification.

Example 5.2: Consider the Poisson regression model where, for a scalar count variable Y_t , it is assumed that Y_t conditional on X_t has a Poisson distribution with mean $E(Y_t|X_t) = \exp(W_t\theta_0)$. As noted in Section 3, the Hausman test is inconsistent against alternatives for which the mean is correctly specified but the distribution is otherwise misspecified. A reasonable approach is to compare the conditional mean and another estimate of the conditional variance. In this case there are no nuisance parameters and $\nabla c_t(m_t(\theta)) \equiv \exp(-W_t\theta)$. Therefore, $\nabla \lambda_t(\theta) \equiv \exp(W_t\theta)W_t$ and $\nabla c_t(\theta)\nabla m_t(\theta) = W_t$. Let ζ_t be a row vector of nonredundant elements of $W_t'W_t$, let $\hat{\theta}_T$ be the Poisson QMLE and let $\hat{U}_t \equiv Y_t - \exp(W_t\hat{\theta}_T)$. The testing procedure is

(i) Perform the multivariate regression of ζ_t on $\exp(W_t\hat{\theta}_T)W_t$, $t=1, \dots, T$ and keep the residuals, say $\ddot{\zeta}_t$.

(ii) Run the OLS regression

$$1 \quad \text{on} \quad (\hat{U}_t^2 - \exp(W_t\hat{\theta}_T))\ddot{\zeta}_t$$

and use TR^2 as asymptotically χ_Q^2 under H_0 where Q is the dimension of ζ_t . ■

In Examples 5.1 and 5.2, the test statistics essentially check whether the information equality holds for the vector of parameters θ_0 , and is not consistent against alternatives for which this equality holds but the distribution is otherwise misspecified. If in Example 5.1 the regression function is correctly specified and the conditional variance of Y_t given X_t is constant, then the

procedure outlined above can not be expected to detect other departures from conditional normality, such as skewness or kurtosis. In Example 5.2, the equality of the conditional mean and conditional variance is being checked. Other departures from the Poisson distribution might not be detected (nor would they be very interesting).

Before ending this section, it is useful to note that Theorem 5.1 is applicable to much more general specification tests for conditional variances, such as the Lagrange Multiplier tests in Breusch and Pagan (1979) and the ARCH tests of Engle (1982). One merely replaces the indicator

$$[(\nabla_{\mathbf{c}_t}(\theta, \pi) \nabla_{\mathbf{m}_t}(\theta, \pi)) \otimes (\nabla_{\mathbf{c}_t}(\theta, \pi) \nabla_{\mathbf{m}_t}(\theta, \pi))]$$

by whatever is desired, as long as the indicator depends only on the predetermined variables X_t and parameters (θ, π) . The result is tests for heteroskedasticity which are robust to departures from the distributional assumption.

EXAMPLE 5.3: Let Y_t be a scalar, and suppose the null hypothesis is

$$H_0: E(Y_t | X_t) = m_t(X_t; \theta_0), \quad \text{some } \theta_0 \in \Theta$$

$$V(Y_t | X_t) = \sigma_0^2 \quad \text{some } \sigma_0^2 > 0 \quad t=1, 2, \dots$$

Let $\hat{\theta}_T$ be the NLLS estimator of θ_0 , and let $\hat{\sigma}_T^2$ be the usual estimator of σ_0^2 based on the sum of squared residuals. The LM test for Q th order ARCH is based upon

$$\sum_{t=Q+1}^T (\hat{U}_t - \hat{\sigma}_T^2) \hat{U}_{t-j}^2 \quad j=1, \dots, Q.$$

The usual LM statistic is $(T-Q)R^2$ from the regression

$$(5.2) \quad \hat{U}_t^2 \quad \text{on} \quad 1, \hat{U}_{t-1}^2, \dots, \hat{U}_{t-Q}^2 \quad t=Q+1, \dots, T.$$

Because $x_t(\theta, \pi) = \pi = \sigma^2$, the statistic that is derived from Theorem 5.1 is $(T-Q)R^2$ from the regression

$$(5.3) \quad 1 \text{ on } (\hat{U}_t^2 - \hat{\sigma}_T^2)(\hat{U}_{t-1}^2 - \hat{\sigma}_T^2), \dots, (\hat{U}_t^2 - \hat{\sigma}_T^2)(\hat{U}_{t-Q}^2 - \hat{\sigma}_T^2) \quad t=Q+1, \dots, T.$$

The regression based form in (5.3) is robust to departures from the conditional normality assumption, and from any other auxiliary assumptions, such as constant conditional fourth moment for U_t^0 . This is to be contrasted with the test derived from (5.2). ■

6. CONCLUSIONS

This paper has developed a general class of specification tests for dynamic multivariate models which impose under H_0 only the hypotheses being tested (correctness of the conditional mean or correctness of the conditional mean and conditional variance). The computationally simple methods proposed here should remove some of the barriers to using robust test statistics in practice.

The general approach used here has several other applications. In particular, the QMLE $\hat{\theta}_T$ can be replaced by any \sqrt{T} -consistent estimator. This is useful in situations where the conditional mean parameters are estimated using a method different than QMLE. An example is a log-linear regression model: let $\theta \equiv (\beta, \sigma^2)$, and

$$(6.1) \quad \log Y_t | X_t \sim N(X_t \beta_0, \sigma_0^2),$$

so that

$$(6.2) \quad E(Y_t | X_t) = \exp(\sigma_0^2/2 + X_t \beta_0).$$

It is easy to estimate θ_0 by MLE in this case since (6.1) suggests OLS of $\log Y_t$ on X_t . Because we are ultimately interested $E(Y_t | X_t)$, QMLE in this example corresponds to NLLS of Y_t on $\exp(X_t \beta)$ (provided

that X_t contains a one). When comparing the log-linear specification to a linear-linear model $E(Y_t|X_t) = X_t\delta_0$, it is useful to use expression (6.2). The robust Davidson-MacKinnon test derived in Section 4 is immediately applicable to the functions $\exp(\hat{\sigma}_T^2/2 + X_t\hat{\beta}_T)$ and $X_t\hat{\delta}_T$ (no matter which model is taken to be the null), where all estimates are obtainable from OLS regressions.

The approach used in this paper also seems to generalize to models that jointly parameterize the conditional mean and variance and are estimated by QMLE using a conditional normality assumption. The multivariate ARCH-in-mean models of the type used by Bollerslev, Engle and Wooldridge (1988) fall into this class. Having robust Lagrange Multiplier tests for these models would allow specification testing of the conditional mean and variance without taking the normality assumption seriously. This research is currently progress.

MATHEMATICAL APPENDIX

For convenience, I include a lemma which is used repeatedly in the proofs of Theorems 3.1 and 5.1.

Lemma A.1: Assume that the sequence of random functions $\{Q_T(W_T, \theta) : \theta \in \Theta, T=1, 2, \dots\}$, where $Q_T(W_T, \cdot)$ is continuous on Θ and Θ is a compact subset of \mathbb{R}^P , and the sequence of nonrandom functions $\{\bar{Q}_T(\theta) : \theta \in \Theta, T=1, 2, \dots\}$, satisfy the following conditions:

$$(i) \sup_{\theta \in \Theta} |Q_T(W_T, \theta) - \bar{Q}_T(\theta)| \xrightarrow{P} 0;$$

(ii) $\{\bar{Q}_T(\theta) : \theta \in \Theta, T=1, 2, \dots\}$ is continuous on Θ uniformly in T .

Let $\ddot{\theta}_T$ be a sequence of random vectors such that $\ddot{\theta}_T - \theta_T^0 \xrightarrow{P} 0$ where $\{\theta_T^0\} \subset \Theta$. Then

$$Q_T(W_T, \ddot{\theta}_T) - \bar{Q}_T(\theta_T^0) \xrightarrow{P} 0.$$

Proof: see Wooldridge (1986, Lemma A.1, p.229). ■

A definition simplifies the statement of the conditions.

Definition A.1: A sequence of random functions $\{q_t(Y_t, X_t, \theta) : \theta \in \Theta, t=1, 2, \dots\}$, where $q_t(Y_t, X_t, \cdot)$ is continuous on Θ and Θ is a compact subset of \mathbb{R}^P , is said to satisfy the Uniform Weak Law of Large Numbers (UWLLN) and Uniform Continuity (UC) conditions provided that

$$(i) \sup_{\theta \in \Theta} \left| T^{-1} \sum_{t=1}^T q_t(Y_t, X_t, \theta) - E[q_t(Y_t, X_t, \theta)] \right| \xrightarrow{P} 0$$

and

$$(ii) \left\{ T^{-1} \sum_{t=1}^T E[q_t(Y_t, X_t, \theta)] : \theta \in \Theta, T=1, 2, \dots \right\} \text{ is}$$

continuous on Θ uniformly in T . ■

In the statement of the conditions, the dependence of functions on the predetermined variables x_t is frequently suppressed for notational convenience. Also, $c_t(\theta, \pi)$ is used as a shorthand for $c_t(m_t(\theta), \eta_t(\pi))$. Similarly for $a_t(\theta, \pi)$. If a gradient operator is not subscripted, the derivative is with respect to all parametric arguments of the function (either θ or (θ, π)). If $a(\theta)$ is a $1 \times L$ vector of the $P \times 1$ vector θ then, by convention, $\nabla_{\theta} a(\theta)$ is the $L \times P$ matrix $\nabla_{\theta} [a(\theta)']$. If $A(\theta)$ is a $Q \times L$ matrix function of the $P \times 1$ vector θ , the matrix $\nabla_{\theta} A(\theta)$ is the $QL \times P$ matrix defined as

$$\nabla_{\theta} A(\theta)' \equiv [\nabla_{\theta} A_1(\theta)' : \dots : \nabla_{\theta} A_Q(\theta)']$$

where $A_j(\theta)$ is the j th row of $A(\theta)$ and $\nabla_{\theta} A_j(\theta)$ is the $L \times P$ gradient of $A_j(\theta)$ as defined above. For simplicity, define

$$\nabla_m^2 c(m, \eta) \equiv \nabla_m [\nabla_m c(m, \eta)']$$

$$\nabla_{m\eta}^2 c(m, \eta) \equiv \nabla_{\eta} [\nabla_m c(m, \eta)']$$

$$\nabla_{\theta}^2 m_t(\theta) \equiv \nabla_{\theta} [\nabla_{\theta} m_t(\theta)']$$

Note that $\nabla_m^2 c(m, \eta)$ is $K^2 \times K$, $\nabla_{m\eta}^2 c(m, \eta)$ is $K^2 \times J$, and $\nabla_{\theta}^2 m_t(\theta)$ is $KP \times P$.

Conditions A.1:

- (i) Θ and Π are compact with nonempty interior;
- (ii) $\{m_t(x_t, \theta) : x_t \in \mathbb{R}^{L+t(L+1)}, \theta \in \Theta\}$ is a sequence of real-valued functions such that $m_t(\cdot, \theta)$ is Borel measurable for each $\theta \in \Theta$ and $m_t(x_t, \cdot)$ is twice continuously differentiable on the interior of Θ for all x_t , $t=1, 2, \dots$;
- (iii) The functions $a: \mathcal{M} \times \mathcal{N} \rightarrow \mathbb{R}$ and $c: \mathcal{M} \times \mathcal{N} \rightarrow \mathbb{R}^K$ are such that

(a) a is continuous on $\mathcal{M} \times \mathcal{N}$ and for each $\eta \in \mathcal{N}$, $a(\cdot; \eta)$ is continuously differentiable on the interior of \mathcal{M} ;

(b) c is twice continuously differentiable on the interior of $\mathcal{M} \times \mathcal{N}$;

(c) For all $(m, \eta) \in \text{interior } \mathcal{M} \times \mathcal{N}$, $m \nabla_m c(m, \eta) = -\nabla_m a(m, \eta)$;

(iv) $T^{-1/2}(\hat{\pi}_T - \pi_T^0) = O_p(1)$;

(v) (a) $\{a_t(\theta, \pi) + m_t(\theta_0)c_t(\theta, \pi)\}$ and $\{U_t^0 c_t(\theta, \pi)\}$

satisfy the WULLN and UC conditions;

(c) θ_0 is the identifiably unique maximizer (see Bates and White (1985)) of

$$T^{-1} \sum_{t=1}^T E[a_t(\theta, \pi_T^0) + m_t(\theta_0)c_t(\theta, \pi_T^0)];$$

(vi) θ_0 is in the interior of Θ , and $\{\pi_T^0\}$ is in the interior of Π uniformly in T ;

(vii) (a) $\{\nabla_{\theta} m_t(\theta)' \nabla_m c_t(\theta, \pi) \nabla_{\theta} m_t(\theta)\}$,
 $\{\nabla_{\theta} m_t(\theta)' \nabla_m^2 c_t(\theta, \pi)' [I_K \otimes U_t(\theta)'] \nabla_{\theta} m_t(\theta)\}$,
 $\{\nabla_{\theta}^2 m_t(\theta)' [I_K \otimes \nabla_m c_t(\theta, \pi) U_t(\theta)']\}$, and
 $\{\nabla_{\pi} \eta_t(\pi)' \nabla_{m\eta}^2 c_t(\theta, \pi)' [I_K \otimes U_t(\theta)'] \nabla_{\theta} m_t(\theta)\}$

satisfy the WULLN and UC conditions.

(b) $A_T^0 \equiv \{T^{-1} \sum_{t=1}^T E[\nabla_{\theta} m_t(\theta_0)' \nabla_m c_t(\theta_0, \pi_T^0) \nabla_{\theta} m_t(\theta_0)]\}$ is $O(1)$

and uniformly positive definite;

(c) $T^{-1/2} \sum_{t=1}^T \nabla_{\theta} m_t(\theta_0)' \nabla_m c_t(\theta_0, \pi_T^0) U_t^0 = O_p(1)$;

(viii) (a) $\{\nabla_{\theta} m_t(\theta)' \nabla_m^2 c_t(\theta, \pi)' [I_K \otimes U_t(\theta)'] \nabla_{\theta} \Lambda_t(\theta, \pi)\}$,
 $\{\nabla_{\pi} \eta_t(\pi)' \nabla_{m\eta}^2 c_t(\theta, \pi) [I_K \otimes U_t(\theta)'] \nabla_{\theta} \Lambda_t(\theta, \pi)\}$,
 $\{\nabla_{\theta} \Lambda_t(\theta, \pi)' [I_K \otimes \nabla_m c_t(\theta, \pi) U_t(\theta)']\}$, and
 $\{\nabla_{\pi} \Lambda_t(\theta, \pi)' [I_K \otimes \nabla_m c_t(\theta, \pi) U_t(\theta)']\}$

satisfy the WULLN and UC conditions.

(ix) (a)

$$\Xi_T^0 \equiv T^{-1} \sum_{t=1}^T E[(\Lambda_t^0 - \nabla_{\theta}^m \Lambda_t^0)' \nabla_m c_t^0 U_t^0, U_t^0 \nabla_m c_t^0 (\Lambda_t^0 - \nabla_{\theta}^m \Lambda_t^0)]$$

is $O(1)$ and uniformly positive definite;

$$(b) \quad \Xi_T^{0-1/2} T^{-1/2} \sum_{t=1}^T (\Lambda_t^0 - \nabla_{\theta}^m \Lambda_t^0)' \nabla_m c_t^0 U_t^0 \stackrel{d}{\rightarrow} N(0, I_Q);$$

$$(c) \quad \{\Lambda_t(\theta, \pi)' \nabla_m c_t(\theta, \pi) U_t(\theta)' U_t(\theta) \nabla_m c_t(\theta, \pi) \Lambda_t(\theta, \pi)\}, \\ \{\Lambda_t(\theta, \pi)' \nabla_m c_t(\theta, \pi) U_t(\theta)' U_t(\theta) \nabla_m c_t(\theta, \pi) \nabla_{\theta}^m \Lambda_t(\theta)\}, \text{ and} \\ \{\nabla_{\theta}^m \Lambda_t(\theta)' \nabla_m c_t(\theta, \pi) U_t(\theta)' U_t(\theta) \nabla_m c_t(\theta, \pi) \nabla_{\theta}^m \Lambda_t(\theta)\}$$

satisfy the UWLLN and UC conditions. ■

Proof of Theorem 3.1: The major task of the proof is establishing the validity of equation (3.24). For notational simplicity, we explicitly consider the case $K = 1$; the case $K > 1$ is similar but notationally cumbersome.

By a weak consistency analog of Bates and White (1985, Theorem 2.2), assumptions (i), (ii), (iii), (iv) and (v) imply that $\hat{\theta}_T \xrightarrow{P} \theta_0$. Consistency of $\hat{\theta}_T$ and (vi) imply that

$$(a.1) \quad P \left[\sum_{t=1}^T \nabla_{\theta}^m \Lambda_t(\hat{\theta}_T) \nabla_m c_t(\hat{\theta}_T) \hat{U}_t = 0 \right] \rightarrow 1 \text{ as } T \rightarrow \infty.$$

Expanding the score $S_T(\hat{\theta}_T, \hat{\pi}_T)$ in a mean value expansion about θ_0 (Jennrich (1969, Lemma 3)) yields

$$(a.2) \quad S_T(\hat{\theta}_T, \hat{\pi}_T) = S_T(\theta_0, \hat{\pi}_T) + \ddot{H}_T(\hat{\theta}_T - \theta_0)$$

where \ddot{H}_T is the hessian with respect to θ with rows evaluating at mean values on the line segment connecting $\hat{\theta}_T$ and θ_0 . For any $\theta \in \text{int } \Theta$,

$$h_t(\theta, \pi) = -\nabla m_t(\theta)' \nabla_m c_t(\theta, \pi) \nabla m_t(\theta) + U_t(\theta) \{ \nabla m_t(\theta)' \nabla_m^2 c_t(\theta, \pi) \nabla m_t(\theta) + \nabla_m c_t(\theta, \pi) \nabla^2 m_t(\theta) \}.$$

Because $\ddot{\Theta}_T^{(j)} \xrightarrow{p} \Theta_0$, $\hat{\pi}_T - \pi_T^0 \xrightarrow{p} 0$, and all components of $h_t(\theta, \pi)$ satisfy the UWLLN and UC conditions by (vii.a), it follows that

$$\ddot{H}_T/T + A_T^0 \xrightarrow{p} 0.$$

By (vii.b), A_T^0 is $O(1)$ and uniformly p.d. Therefore, \ddot{H}_T is nonsingular with probability approaching one, so by Lemma A.1, $(\ddot{H}_T/T)^{-1} + A_T^{0-1} \xrightarrow{p} 0$. Combined with (a.1) and (a.2) this implies that

$$(a.3) \quad T^{1/2}(\hat{\Theta}_T - \Theta_0) = A_T^{0-1} T^{-1/2} S_T(\Theta_0, \hat{\pi}_T) + o_p(1).$$

Next, note that by a mean value expansion about π_T^0 ,

$$T^{-1/2} S_T(\Theta_0, \hat{\pi}_T) = T^{-1/2} S_T(\Theta_0, \pi_T^0) + [\nabla_{\pi} \ddot{S}_T/T] T^{1/2}(\hat{\pi}_T - \pi_T^0) + o_p(1)$$

where $\nabla_{\pi} \ddot{S}_T$ has rows evaluated at $(\Theta_0, \ddot{\pi}_T^{(j)})$ and $\ddot{\pi}_T^{(j)}$ are mean values between $\hat{\pi}_T$ and π_T^0 . By (vii.a), $\nabla_{\pi} s_t(\theta, \pi)$ satisfies the UWLLN and UC conditions so that by (iv) and Lemma A.1, $[\nabla_{\pi} \ddot{S}_T/T] - E[\nabla_{\pi} S_T(\Theta_0, \pi_T^0)/T] = o_p(1)$. But, under H_0 , $E[\nabla_{\pi} s_t(\theta, \pi_T^0) | X_t] = 0$, so that $\nabla_{\pi} \ddot{S}_T/T = o_p(1)$. Combined with (iv) this shows that

$$(a.4) \quad T^{-1/2} S_T(\Theta_0, \hat{\pi}_T) = T^{-1/2} S_T(\Theta_0, \pi_T^0) + o_p(1).$$

Substituting (a.4) into (a.3) gives the standard asymptotic equivalence for the QMLE,

$$T^{1/2}(\hat{\Theta}_T - \Theta_0) = A_T^{0-1} T^{-1/2} S_T^0 + o_p(1);$$

in particular, $T^{1/2}(\hat{\Theta}_T - \Theta_0) = O_p(1)$ by (vii.c).

Next, consider equation (3.24). First, (vii:b) guarantees that B_T^0 exists for sufficiently large T and is $O(1)$. Rewriting (3.24) gives

$$(a.6) \quad \hat{\theta}_T = T^{-1/2} \sum_{t=1}^T (\nabla_{m_t} \hat{c}_t^{1/2} \hat{\Lambda}_t - \nabla_{m_t} \hat{c}_t^{1/2} \nabla_{\theta_t} \hat{m}_t^0 B_T^0)' \nabla_{m_t} \hat{c}_t^{1/2} \hat{U}_t' \\ - T^{1/2} (\hat{B}_T - B_T^0)' T^{-1} \sum_{t=1}^T \nabla_{\theta_t} \hat{m}_t' \nabla_{m_t} \hat{c}_t \hat{U}_t'.$$

By (a.1), the second term on the right hand side of (a.6) is zero with probability approaching one. Taking a mean value expansion about (θ_0, π_0^0) of the first term on the right hand side of (a.6) and applying Lemma A.1, (vii), and (viii) yields

$$(a.7) \quad \hat{\theta}_T = T^{-1/2} \sum_{t=1}^T (\Lambda_t^0 - \nabla_{\theta_t} m_t^0 B_T^0)' \nabla_{m_t} c_t^0 U_t^0 \\ + \left[T^{-1} \sum_{t=1}^T [(\nabla_{\theta_t} \Lambda_t^0 - \nabla_{\theta_t}^2 m_t^0 B_T^0)' \nabla_{m_t} c_t^0 U_t^0] \right] T^{1/2} (\hat{\theta}_T - \theta_0) \\ + \left[T^{-1} \sum_{t=1}^T [(\Lambda_t^0 - \nabla_{\theta_t} m_t^0 B_T^0)' \nabla_{m_t}^2 c_t^0 \nabla_{\theta_t} m_t^0 U_t^0] \right] T^{1/2} (\hat{\theta}_T - \theta_0) \\ - \left[T^{-1} \sum_{t=1}^T [(\Lambda_t^0 - \nabla_{\theta_t} m_t^0 B_T^0)' \nabla_{m_t} c_t^0 \nabla_{m_t}^0] \right] T^{1/2} (\hat{\theta}_T - \theta_0) \\ + \left[T^{-1} \sum_{t=1}^T [\nabla_{\pi_t} \Lambda_t^0, \nabla_{m_t} c_t^0 U_t^0] \right] T^{1/2} (\hat{\pi}_T - \pi_0) \\ + \left[T^{-1} \sum_{t=1}^T [\Lambda_t^0, \nabla_{m_t \eta_t}^2 c_t^0 \nabla_{\pi_t} \eta_t^0 U_t^0] \right] T^{1/2} (\hat{\pi}_T - \pi_0) + o_p(1).$$

Under H_0 , $E[U_t^0 | X_t] = 0$, so that the first term in each of lines two, three, five, and six of (a.7) has zero expectation by the law of iterated expectations. Because each of these terms satisfies the WLLN, $T^{1/2} (\hat{\theta}_T - \theta_0) = o_p(1)$, and $T^{1/2} (\hat{\pi}_T - \pi_0) = o_p(1)$, the expressions in lines two, three, five, and six are all $o_p(1)$. By definition of B_T^0 ,

$$T^{-1} \sum_{t=1}^T E[(\Lambda_t^0 - \nabla_{\theta_t} m_t^0 B_T^0)' \nabla_{m_t} c_t^0 \nabla_{m_t}^0] = 0$$

so that the term in line four is also $o_p(1)$. This establishes that

$$\hat{\vartheta}_T = T^{-1/2} \sum_{t=1}^T (\Lambda_t^0 - \nabla_{\Theta} m_t^0 B_T^0)' \nabla_m c_t^0 U_t^0 + o_p(1)$$

so that (3.24) in the text holds. By (ix.a), Ξ_T^0 is $O(1)$ and uniformly p.d. By (ix.b), $\Xi_T^{0-1/2} \hat{\vartheta}_T \xrightarrow{d} N(0, I_Q)$. Condition (ix.c) ensures that $\hat{\Xi}_T$ is a consistent estimator of Ξ_T^0 , and consequently is positive definite with probability approaching one. Therefore, $\hat{\vartheta}_T' \hat{\Xi}_T^{-1} \hat{\vartheta}_T \xrightarrow{d} \chi_Q^2$, and this completes the proof. ■

In what follows, let

$$v_t(\theta) \equiv [\text{vec } U_t(\theta)' U_t(\theta)]'$$

$$x_t(\theta, \pi) \equiv [\text{vec } \nabla_m c_t(\theta, \pi)^{-1}]'$$

$$\Lambda_t(\theta, \pi) \equiv [(\nabla c_t(\theta, \pi) \nabla m_t(\theta)) \otimes (\nabla c_t(\theta, \pi) \nabla m_t(\theta))].$$

Conditions B.1: Conditions (i)-(vi) in A.1 hold. In addition,

(vii') The following functions satisfy the WULLN and UC

conditions:

$$\{\nabla v_t(\theta) \Lambda_t(\theta, \pi)\}, \{\nabla x_t(\theta, \pi) \Lambda_t(\theta, \pi)\}, \{\nabla v_t(\theta) \nabla x_t(\theta, \pi)'\},$$

$$\{\nabla x_t(\theta, \pi)' \nabla x_t(\theta, \pi)\}, \{\nabla x_t(\theta, \pi)' \Lambda_t(\theta, \pi)\},$$

$$\{\nabla \Lambda_t(\theta, \pi)' [I_{P+N} \otimes v_t(\theta)']\}, \{\nabla \Lambda_t(\theta, \pi)' [I_{P+N} \otimes x_t(\theta)']\},$$

$$\{\nabla^2 x_t(\theta, \pi)' [I_{P+N} \otimes v_t(\theta)']\}, \{\nabla^2 x_t(\theta, \pi)' [I_{P+N} \otimes x_t(\theta)']\};$$

$$(viii') \quad T^{-1/2} \sum [v_t(\theta_0) - x_t(\theta_0, \pi_0)]' \nabla x_t(\theta_0, \pi_0) = O_p(1);$$

$$(ix') \quad (a) \quad \Xi_T^0 \equiv T^{-1} \sum_{t=1}^T E[\psi_t(\theta_0, \pi_0, \Gamma_T^0)' \psi_t(\theta_0, \pi_0, \Gamma_T^0)] \text{ is } O(1) \text{ and}$$

uniformly p.d.;

$$(b) \quad \Xi_T^{0-1/2} T^{-1/2} \sum_{t=1}^T \psi_t(\theta_0, \pi_0, \Gamma_T^0) \xrightarrow{d} N(0, I_Q);$$

(c) $\{\psi_t(\theta, \pi, \Gamma)' \psi_t(\theta, \pi, \Gamma)\}$ satisfies the WULLN and UC conditions.

Proof of Theorem 5.1: For notational simplicity, again consider the case $K = 1$. First, because (i)-(vi) of A.1 hold, $T^{1/2}(\hat{\theta}_T - \theta_0) = O_p(1)$ and $T^{1/2}(\hat{\pi}_T - \pi_0) = O_p(1)$. Next, note that

$$\begin{aligned}
 \text{(a.8)} \quad & T^{-1/2} \sum_{t=1}^T [U_t(\hat{\theta}_T)^2 - \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] [\Lambda_t(\hat{\theta}_T, \hat{\pi}_T) - \hat{\Gamma}_T \nabla \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] \\
 &= T^{-1/2} \sum_{t=1}^T [U_t(\hat{\theta}_T)^2 - \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] [\Lambda_t(\hat{\theta}_T, \hat{\pi}_T) - \Gamma_T^0 \nabla \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] \\
 &\quad - (\hat{\Gamma}_T - \Gamma_T^0) T^{-1/2} \sum_{t=1}^T [U_t(\hat{\theta}_T)^2 - \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] \nabla \gamma_t(\hat{\theta}_T, \hat{\pi}_T).
 \end{aligned}$$

It is straightforward to show that (vii') and Lemma A.1 imply that $(\hat{\Gamma}_T - \Gamma_T^0) = o_p(1)$. Next, a mean value expansion, (vi'), (vii'), $T^{1/2}(\hat{\theta}_T - \theta_0) = O_p(1)$ and $T^{1/2}(\hat{\pi}_T - \pi_0) = O_p(1)$ imply that

$$T^{-1/2} \sum_{t=1}^T [U_t(\hat{\theta}_T)^2 - \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] \nabla \gamma_t(\hat{\theta}_T, \hat{\pi}_T) = O_p(1).$$

Expanding the first term on the right hand side of (a.8) about (θ_0, π_0) yields

$$\begin{aligned}
 & T^{-1/2} \sum_{t=1}^T [U_t(\hat{\theta}_T)^2 - \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] [\Lambda_t(\hat{\theta}_T, \hat{\pi}_T) - \Gamma_T^0 \nabla \gamma_t(\hat{\theta}_T, \hat{\pi}_T)] \\
 &= T^{-1/2} \sum_{t=1}^T [(U_t^0)^2 - \gamma_t^0] [\Lambda_t^0 - \Gamma_T^0 \nabla \gamma_t^0] \\
 &\quad + T^{-1} \sum_{t=1}^T [(U_t^0)^2 - \gamma_t^0] [\nabla \Lambda_t^0 - \Gamma_T^0 \nabla^2 \gamma_t^0] \cdot T^{1/2}(\hat{\delta} - \delta_0) \\
 &\quad - T^{-1} \sum_{t=1}^T 2 \nabla m_t^0 \cdot U_t^0 [\Lambda_t^0 - \Gamma_T^0 \nabla \gamma_t^0] \cdot T^{1/2}(\hat{\theta}_T - \theta_0) \\
 &\quad - T^{-1} \sum_{t=1}^T [\Lambda_t^0 - \Gamma_T^0 \nabla \gamma_t^0] \cdot \nabla \gamma_t^0 T^{1/2}(\hat{\delta} - \delta_0) + o_p(1)
 \end{aligned}$$

where $\delta \equiv (\theta', \pi')'$. Under H_0 , $E[U_t^0 | X_t] = 0$ and $E[(U_t^0)^2 | X_t] = \gamma_t^0$; therefore, the second and third terms on the RHS of (a.8) are $o_p(1)$. Γ_T^0 is defined so that

$$T^{-1} \sum_{t=1}^T E[\nabla \gamma_t^0 (\Lambda_t^0 - \Gamma_T^0 \nabla \gamma_t^0)] = 0.$$

Therefore, the fourth term is $o_p(1)$. We have shown that

$$(a.9) \quad T^{-1/2} \sum_{t=1}^T [\hat{U}_t^2 - \hat{\gamma}_t] [\hat{\Lambda}_t - \hat{\Gamma}_T \nabla \hat{\gamma}_t] = \\ T^{-1/2} \sum_{t=1}^T [U_t^{o2} - \gamma_t^0] [\Lambda_t^0 - \Gamma_T^0 \nabla \gamma_t^0] + o_p(1).$$

Recalling that ψ_t^0 denotes the nonredundant elements in the summation on the right hand side, it follows from (ix'.a,b) that

$$\Xi_T^{-1/2} T^{-1/2} \sum_{t=1}^T \psi_t^0 \stackrel{d}{\rightarrow} N(0, I_Q). \text{ Combined with (ix'c), this shows that}$$

$$\hat{\xi}_T \stackrel{d}{\rightarrow} \chi_Q^2 \text{ under } H_0 \text{ and completes the proof. } \blacksquare$$

References

- Andrews, D.W.K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465-1471.
- Bates, C., and H. White (1985): "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory*, 1, 151-175.
- Bollerslev, T., R.F. Engle, and J.M. Wooldridge (1988): "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy* 96, 116-131.
- Breusch, T.S., and A.R. Pagan (1979): "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.
- Cox, D.R. (1961): "Tests of Separate Families of Hypotheses," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press, 105-123.
- Cox, D.R. (1962): "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society, Series B*, 24, 406-424.
- Davidson, R., and J.G. MacKinnon (1981): "Several Tests of Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49, 781-793.
- Domowitz I., and H. White (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20, 35-58.
- Engle, R. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation," *Econometrica*, 50, 987-1008.
- Gallant, A.R., and H. White (1986): "A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models," UCSD Department of Economics Discussion Paper.
- Gourieroux, C., A. Monfort, and A. Trognon (1984a): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681-700.
- Gourieroux, C., A. Monfort, and A. Trognon (1984b): "Pseudo Maximum Likelihood Methods: Applications," *Econometrica*, 52, 701-720.
- Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

- Hausman, J.A., Ostrom, B.D., and D.A. Wise (1984): "Air Pollution and Lost Work," NBER Working Paper 1263.
- Hsieh, D.A. (1983): "A Heteroskedasticity-Consistent Covariance Matrix Estimator for Time Series Regressions," *Journal of Econometrics*, 22, 281-290.
- Jennrich, R. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633-643.
- Mizon, G., and J.-F. Richard (1986): "The Encompassing Principle and Its Application to Testing Nonnested Hypotheses," *Econometrica*, 54, 657-678.
- Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.
- Newey, W. (1987): "Expository Notes on Uniform Laws of Large Numbers," mimeo, Princeton University.
- Tauchien, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.
- White, H. (1980a): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.
- White, H. (1980b): "Nonlinear Regression on Cross Section Data," *Econometrica*, 48, 721-746.
- White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-26.
- White, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.
- White, H. (1985a): "Specification Testing in Dynamic Models," Invited Paper at the Fifth World Congress of the Econometric Society, Cambridge, Massachusetts, August 1985.
- White, H. (1985b): "Estimation, Inference, and Specification Testing," manuscript, UCSD Department of Economics.
- White, H., and I. Domowitz (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143-162.
- Wooldridge, J.M. (1986): "Asymptotic Properties of Econometric Estimators," UCSD Ph.D. dissertation.
- Wooldridge, J.M. (1987): "On the Existence of Conditional Densities in Econometric Models," mimeo, MIT.

Date Due

Vols

AG 2 '89
NOV 12 1989
NOV 21 1990
MAY 17 1991
SEP. 15 1993

MIT LIBRARIES



3 9080 005 336 547

